Expert Systems with Applications 39 (2012) 9166–9180

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Lexicon-based Comments-oriented News Sentiment Analyzer system

A. Moreo *, M. Romero, J.L. Castro, J.M. Zurita

Dept. of Computer Science and Artificial Intelligence, University of Granada, Spain

ARTICLE INFO

Keywords:
Sentiment Analysis
Lexicon-based
News analysis
Feature mining
Focus detection

ABSTRACT

Thanks to the technological revolution that has accompanied the Web 2.0, users are able to interact intensively on the Internet, as reflected in social networks, blogs, forums, etc. In these scenarios, users can speak freely on any relevant topic. However, the high volume of user-generated content makes a manual analysis of this discourse unviable. Consequently, automatic analysis techniques are needed to extract the opinions expressed in users' comments, given that these opinions are an implicit barometer of unquestionable interest for a wide variety of companies, agencies, and organisms. We thus propose a lexicon-based Comments-oriented News Sentiment Analyzer (LCN-SA), which is able to deal with the following: (a) the tendency of many users to express their views in non-standard language; (b) the detection of the target of users' opinions in a multi-domain scenario; (c) the design of a linguistic modularized knowledge model with low-cost adaptability. The system proposed consists of an automatic Focus Detection Module and a Sentiment Analysis Module capable of assessing user opinions of topics in news items. These modules use a taxonomy-lexicon specifically designed for news analysis. Experiments show that the results obtained thus far are extremely promising.

1. Introduction

The Internet is currently evolving towards the Web 2.0. This trend as well as the fact that the Internet is now within almost everyone's reach because of cheaper hardware has led to a cultural revolution. People from all over the world are now able to interact with each other. This has generated social networks, blogs, forums, etc., where they can enjoy both freedom of speech and easy access to all types of information. Within such contexts, users are able to express their opinions on any relevant topic. This type of scenario includes the interactive press or online news sites with publications on current events. These sites encourage user communities to say what they think about breaking news in categories ranging from sports to controversial social debates. The proliferation of user comments, favored by anonymity, generates large quantities of unstructured information. This knowledge is known as user-generated content, and its analysis helps to shape a social barometer pertaining to any issue. The study of this barometer provides a popularity measurement of news in a global context, based on user comments. For example, a headline concerning a new economic measure not only generates evaluations of the measure itself, but also of the government that enacted it. Nonetheless, given the sheer quantity of news published on the Internet, it is very difficult, if not impossible, to manually analyze it all. An automatic method is thus needed that is capable of processing and analyzing the information conveyed in news comments. This article describes a new approach to the automated analysis of user-generated content.

The computational study of opinions, sentiments, and emotions expressed in texts is known as Sentiment Analysis or Opinion Mining (Chen & Zimbra, 2010; Pang & Lee, 2008). This includes the automatic extraction of opinions and the analysis of sentiment. Although opinion is very broad concept, Sentiment Analysis has thus far mainly focused on positive and negative sentiments. This research area evaluates words and sentences in opinion-expressing documents with a view to studying their **subjectivity**, **polarity**, and **strength**: (i) subjectivity is the extent to which a text is objective, and whether it contains sentiment expressions with subjective views; (ii) polarity is the extent to which the text expresses a positive or negative sentiment; (iii) strength is the degree of polarity or intensity of the opinion. Furthermore, another objective is the discovery of the main topics on which user opinions are expressed.

Earlier studies on Sentiment Analysis have focused on the classification of product reviews (Denecke, 2008; Hu & Liu, 2004a; Miao, Li, & Dai, 2008; Mullen & Collier, 2004; Pang, Lee, & Vaithyanathan, 2002). Their goal is the extraction of positive or negative sentiments in user opinions of a product or some of its features in order to classify the reviews of the product. This task, which is generally known as **overall sentiment**, operates on a document level. Therefore, the sentiment extracted is atomically attached to the reviewed product. This type of evaluation mostly centers on polarity

* Corresponding author. Address: C/Periodista Daniel Saucedo Aranda s/n, E-18071 Granada, Spain. Tel.: +34 958244019; fax: +34 958243317.

E-mail addresses: moreo@decsai.ugr.es (A. Moreo), manudb@decsai.ugr.es (M. Romero), castro@decsai.ugr.es (J.L. Castro), zurita@decsai.ugr.es (J.M. Zurita).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.057

(positiveness or negativity) and optionally, on its strength or intensity. This research area is certainly useful for companies offering products or services. Instead of conducting costly market studies or customer satisfaction analyses, companies are able to analyze published reviews to determine user affinity to their products. In fact, related research studies have focused on the following: (i) the acceptance of user reviews rather than those in other more conventional information sources (Bickart & Schindler, 2001); (ii) the commercial interest deposited in online user opinions of products and services (Hoffman, 2008); (iii) the growing influence of these views on the purchasing decisions of other users (Chevalier & Mayzlin, 2006).

Studies more closely linked to the context of our problem, such as socio-political studies, show that the "geopolitical" web can improve the extraction of citizens' opinions regarding the most important public issues of debate, particularly political issues (Adamic & Glance, 2005; Malouf & Mullen, 2008). These studies are even able to classify news items based on their evaluation (Benamara, Cesarano, Picariello, Reforgiato, & Subrahmanian, 2007; Jindal & Liu, 2006; Wiebe, Wilson, Bruce, Bell, & Martin, 2004). Nevertheless, the extraction of sentiment from any informative text is very difficult because news items are supposedly objective and free of polarity. Thus, it seems more appropriate to focus on comments in order to analyze opinions of the news.

Most Sentiment Analysis techniques can be divided into machine-learning approaches and dictionary-based approaches. Even though machine-learning approaches have made significant advances (Mullen & Collier, 2004; Pang et al., 2002; Tan & Zhang, 2008; Turney, 2002) in sentiment classification, applying them to news comments require labeled training data sets. The compilation of these training data requires considerable time and effort, especially since data should be current. In order to alleviate this task, applications to generate annotated corpus data were proposed (Wiebe, Wilson, & Cardie, 2005).

In contrast, dictionary-based approaches extract the polarity of each sentence in a document. Afterwards, the sense of the opinion words in the phrase is analyzed in order to group polarities, and thus classify the sentiment of the text. Generally speaking, the techniques that follow this approach are based on lexicons, and use a dictionary of words mapped to their semantic value (Denecke, 2008; Devitt & Ahmad, 2007), such as MPQA lexicon (Wilson, Wiebe, & Hoffmann, 2005), WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), or SentiWordNet (Esuli & Sebastiani, 2006), an enhanced lexical resource for supporting sentiment classification. Although these dictionaries are usually handmade, some approaches use "seed words" to automatically expand knowledge (Godbole, Srinivasaiah, & Skiena, 2007; Hatzivassiloglou & McKeown, 1997). Dictionary-based approaches have important advantages, such as the fact that once they are built, no training data are necessary. Unfortunately, they also have certain drawbacks. First, most are designed as glossaries of general language words, often based on WordNet, and thus they do not contain either technical terms or colloquial expressions (Example 1.a). Secondly, since they are unable to consider context-dependent expressions, such systems achieve very limited accuracy in multi-domain scenarios (Denecke, 2009) because the connotation of certain words can be either positive or negative, depending on the context in which the words are used (Example 1.b).

	Comment	Analysis
1.a	"The prime minister has *lost his mind*"	Negative opinion of the subject "The prime minister", using colloquial language
1.b	"The battery life is *too short*". "It wasn't so bad. The wait was *short*"	The sentiment of the adjective "short" is negative in the first phrase and positive in second one

In addition to overall sentiment, there is another strategy in Sentiment Analysis that also considers the subcomponents and attributes of the product or service individually. Such approaches provide a review of each feature (Ding, Liu, & Yu, 2008; Hu & Liu, 2004a; Jindal & Liu, 2006; Miao et al., 2008), rather than giving a single overall rating. This kind of analysis is known as **feature mining** or **feature-based** Sentiment Analysis (Example 1.c), and is often applied to product reviews. In this case, feature extraction is usually based on the strong assumption that a review is of a single product (Godbole et al., 2007; Hu & Liu, 2004a) or a comparison of several products of the same type (Popescu & Etzioni, 2005), such as comparing the features of different cameras. For this reason, although the term "feature" is appropriate in a product review, this is not necessarily the case in news analysis because there may be more than one object evaluated. Other authors have proposed different terms such as "topic" (Kim & Hovy, 2004) or "aspect" (Kobayashi, Inui, & Matsumoto, 2007). This research study uses "focus", which is more appropriate in this context to refer the entity target of opinion.

Our proposal is that all focuses in the news comments with their independent valuation must be analyzed. Thus, the extraction of features should be multifocal. The same expression could refer to features of different objects depending on the context (Example 1.d). To resolve the ambiguity of these sentences, we propose an automatic method to detect opinion targets. This set of focuses defines the context of the news, and allows us to isolate linguistic interferences of expressions with each other in order to disambiguate them.

	Comment	Analysis
1.c	"The product has *good quality*, however, is *quite expensive*"	Two opposing opinions on product features can be seen for the same product. Polarity analysis would be wrong if only the product itself were considered
1.d	"The *cost* of X has considerably increased the *cost* of the shares of the company Y"	The "cost" is a property of entities X and Y. It is essential to understand them as two different features belonging to different entities

Comment-oriented polarity extraction also provides new challenges for Sentiment Analysis, as there is a high probability that users express their opinions of focuses that do not explicitly appear in the body of the news article. In addition, given that comment holders (people that express opinions) are usually anonymous, this often leads to malicious users expressing offensive opinions, or even using their comments to advertise their own websites. Although some news media sites allow users to denounce such behavior, this is not always the case. Accordingly, useful information is sometimes mixed with noisy data that makes analysis more difficult. Detecting and filtering out irrelevant information in user-generated content is a subtask of vital importance when

performing Sentiment Analysis known as **Opinion Spam Detection** (Jindal & Liu, 2008). Our approach includes a comments filter that discards comments that are likely to be noisy.

Our lexicon-based approach consists of a practical system to deal with the difficulties of the analysis of user comments on news articles. As a starting point, we consider that each comment may convey an opinion on the general topic content in the news item or even on another specific topic related to the general topic. Comments on Example 1.e may be on the following news headline: "The X football team wins the European Championship".

	Comment	Analysis
1.e	"The Y team is much better!" "Player Z is the best of the team"	Both comments convey opinions on focuses different from the main focus. They must be analyzed separately after obtaining the focuses of the comment

We propose a system that performs the following actions. Firstly, comments that could introduce inconclusive information are filtered out in order to make the analysis more reliable. Next, all the explicitly or implicitly commented focuses are automatically recognized through an innovative Opinion Focus Detection algorithm. The valuation of each focus is computed with a taxonomy-lexicon as explained in the following sections. As a result, the algorithm obtains a focus classification, which determines polarity and strength. Finally, mining techniques are applied to generate easily interpretable summaries. Fig. 1 shows an example of the sentiment report computed on a sports news item that will be later explained in more detail.

The structure of our lexicon is specifically designed to analyze news comments. It has been built bearing in mind the cross-domain nature of news items allowing the adaptation of the lexicon.

The rest of this article is organized as follows. An overview of previous work on Sentiment Analysis is presented in Section 2. Section 3 describes the structure of our application, and Section 4 provides a detailed explanation of our lexicon and its structure. The method of analysis is outlined in Section 5. Finally, Section 6 discusses the experimental validation of our method, and Section 7 concludes with a discussion of results and future research.

2. Related works

This section describes the state of the art in Sentiment Analysis, and discusses the most relevant research in the field. First, we

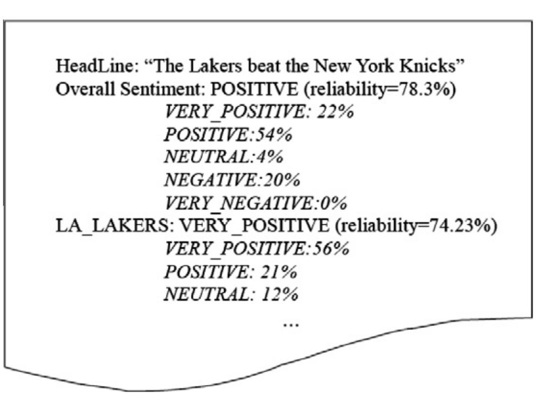

Fig. 1. Fragment of sentiment report.

focus on the user-comment approaches that are most closely related to our research. This is followed by a discussion of machine-learning methods, dictionary-based approaches, and their variants. Finally, we give an overview of Feature Mining, a variant of Sentiment Analysis, which as of late has received a considerable amount of attention from the research community.

In Hu, Sun, and Lim (2007), the authors performed an automatic blog posts summary of the information contained in user comments. They used a graph-based system of weights that draws on the words in the most cited comments related to the most popular topics. Their system then selected the blog post phrases containing the most representative words obtained previously. In (Delort, 2006), Delort extracted clusters of comments. Interesting clusters were selected manually and used to extract the blog sentences most closely related to the comments in the cluster. Agrawal, Rajagopalan, Srikant, and Xu (2003) classified users on opposing sides of an online discussion by means of a graph linking them to comments of the type "answer to". In their study, Malouf and Mullen (2008) ranked users according to their political orientation (left-wing, rightwing, other), based on comments made in American policy forums. For this purpose, the authors used a variation of the PMI-IR method, Naïve Bayes, and a social network analysis method, based on graphs.

Machine-learning and dictionary-based approaches have been applied to product reviews with promising results. Turney developed an unsupervised learning algorithm to classify texts as recommended or not recommended (Turney, 2002). This algorithm, known as Pointwise Mutual Information and Information Retrieval (PMI-IR), calculated the semantic orientation of a sentence by assigning the numerical value resulting from the information shared by the sentence and the word "excellent" minus the information shared by the sentence and the word "poor". The text was classified as *recommended* if the average value was positive, and the magnitude of this numerical value was regarded as indicative of the strength of its semantic orientation. This algorithm has been used in many subsequent investigations. Pang et al. (2002) used three machine-learning techniques to classify IMBD movie reviews as positive or negative. Their conclusion showed that these three techniques (Naïve Bayes, classification maximum entropy, and SVM) outperformed the human-generated baseline, and that SVM was the technique that yielded the best results. In Zhang and Varadarajan (2006), the authors developed several regression models to predict a review's usefulness, on the assumption that this usefulness is orthogonal to its polarity. They concluded that shallow syntactic features were the most influential utility predictors, and remarked that a review's usefulness depended heavily on linguistic style. In their approach, Esuli and Sebastiani (2005) proposed a semi-supervised method of performing a binary classification of texts as positive or negative, assuming that terms with a similar orientation tend to have similar definitions (glosses). Benamara et al. (2007), proposed a linguistic approach to determine the strength and polarity of a topic in a given text, and applied a technique based on adjective and adverb combinations (AACs). They presented three scoring axioms which defined the strength and polarity of a given AAC. Nasukawa and Yi offer an alternative method, and propose extracting sentiments associated with subjects that recur throughout the text, rather than providing a valuation of the document as a whole Nasukawa and Yi (2003). Accordingly, they used a syntactic parser to identify relationships between sentiment expressions and the subjects on which they give an opinion. Afterwards, a sentiment lexicon was used to establish the polarity of the sentiments expressions. Other examples of sub-sentential machine learning methods include fully-supervised work (McDonald, Hannan, Neylon, Wells, & Reynar, 2007) or weakly-supervised models (Täckström & McDonald, 2011; Yessenalina, Yue, & Cardie, 2010).

Within dictionary-based approaches, systems generally use pre-developed dictionaries containing the polarity of words or phrases. The most frequently used resource is currently SentiWordNet, which has been employed in a large number of research studies. In Denecke (2008), the authors use SentiWordNet to determine the polarity of phrases in a multilingual context and classify documents according to polarity. Ohana and Tierney (2009) also used SentiWordNet in a study in which the dataset is a set of film reviews. They performed a similar classification, and concluded that the results provided by SentiWordNet were close to the results obtained with handmade lexicons. Meanwhile, Dang et al. developed an algorithm that combines content-free (lexical, syntactic and structural features), content-specific (keywords and phrases by n-grams) and sentiment techniques (SentiWordNet) for the classification of online product reviews (Dang, Zhang, & Chen, 2010). These researchers concluded that the combination of machine-learning techniques and dictionary-based techniques substantially improved sentiment classification. Godbole et al. (2007) implemented a lexicon-based system for news and blogs analysis built on top of the Lidia text analysis system. They propose a method to expand candidate seed lists opinion words through WordNet. Several automatic techniques to create lexicons have been proposed (Neviarouskaya, Prendinger, & Ishizuka, 2011; Tan & Wu, 2011). However there is no evidence that those lexicons perform better than manually-built ones in cross-domain scenarios (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). In (Wu & Tan, 2011) the authors propose a framework for cross-domain sentiment classification, defining a "bridge" between one existed domain to a target domain. There have been some papers evaluated on the MPQA corpus.[1] Those studies have included lexicon and machine learning based approaches (Wilson et al., 2005) and have tackled structured opinion extraction (Choi, Breck, & Cardie, 2006). There are also approaches that use taxonomies for product Sentiment Analysis based on component-subcomponent scheme (Carenini, Ng, & Zwart, 2005).

Feature Mining is a variant of Sentiment Analysis that also focuses on capturing the particular sentiment evoked by an object. However, it is based on the valuation of its features and subcomponents. Jindal and Liu proposed a comparative sentence-based method (Jindal & Liu, 2006) capable of extracting the set of relations between text entities (e.g. products or product features) by using two types of Sequential Rules. The features of the products were thus identified and specified. The use of this method in product reviews is very interesting because comparisons of products are very common on the Internet.

Other research based on feature mining is that of Miao et al. (2008). In this study, the researchers extracted product features and related opinions from unstructured reviews. Their algorithm received a semi-structured set of reviews with prior knowledge that was gradually enriched with results using linguistic similarity features. Hu and Liu (2004a) used data mining techniques and natural language processing techniques to obtain the feature polarity of a product that had been reviewed. It was assumed that the main features and their valuation explicitly appeared as nouns or noun phrases in the text of the product review. Frequency distributions were used to find the features by proximity. Afterwards, the polarity of the comments on each feature was obtained, and the results summarized, using WordNet.

Ding et al. (2008) considered the problem of context in feature mining. Since the same opinion word can have different orientations in different contexts, both opinion words and features were treated as a tuple called "opinion context". In this way, their system provided a valuation of features based on such tuples by using a lexicon. The system also provided techniques to correctly determine the tuple when ambiguity occurred in opinions related to one feature. In Archak, Ghose, and Ipeirotis (2007), the authors used a hybrid of conventional text-mining techniques and an econometric model (similar to hedonic regression) to estimate the strength and polarity of product features. Hu and Liu (2004b) proposed a set of techniques to detect product features, their frequency of occurrence in the document, and the valuation given by the users. In Yi, Nasukawa, Bunescu, and Niblack (2003), the authors developed a feature term extraction system based on a mixture language model and likelihood ratio. A sentiment lexicon was used to assign sentiment phrases to features. Also, important works on topic models for opinion-topic extraction could be consulted in Lue, Castellanos, Dayal, and Zhai (2011), Titov and McDonald (2008), Li, Feng, Wang, Yu, and He (2008). These use varying levels of supervision, being evaluated on reviews.

3. High-performance system for news analysis

In this section, the general architecture of our high-performance LCN-SA is explained. Our goal was to extract users' opinions by analyzing their comments. We also aim to analyze not only the sentiment in the entire document (**overall sentiment**), but also the sentiment evoked by each discussion topic (**sentiment focus**). Therefore, two major problems should be addressed: (i) the identification of those discussion topics on which users have expressed their opinions; (ii) the extraction of their sentiments in such a way as to avoid interferences between them in the analysis. However, to achieve this, knowledge in the lexicon must be adapted to the news domain being analyzed. Spam comments should also be detected and discarded in order to prevent noisy data interferences.

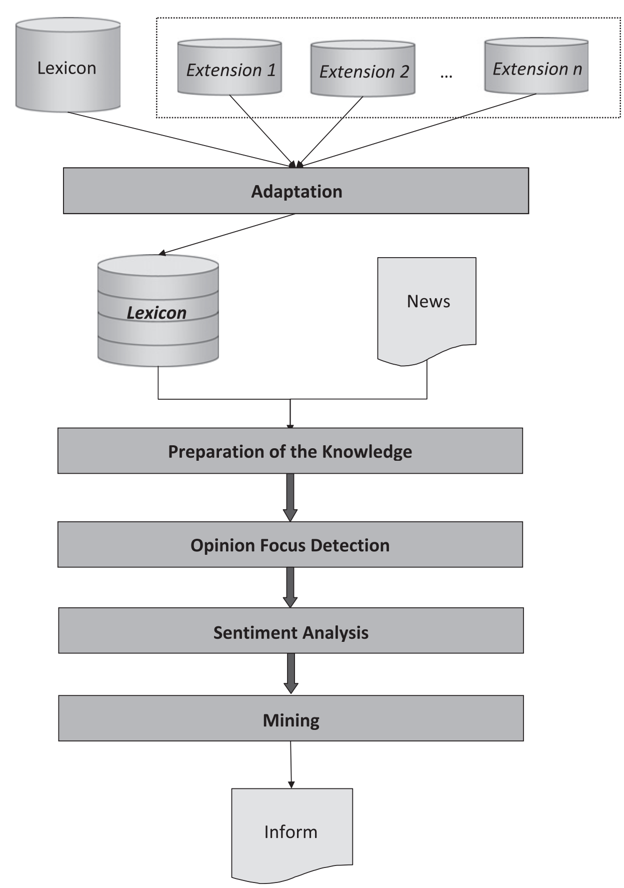

Fig. 2. LCN-SA architecture.

[1] http://www.cs.pitt.edu/mpqa/databaserelease/.

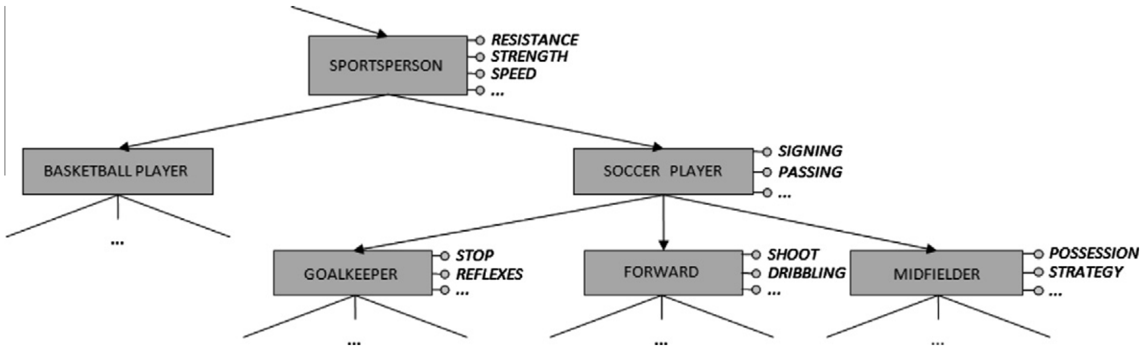

Fig. 3. Section of the lexicon related to SPORTSPERSON and SOCCER_PLAYER objects.

We thus propose a lexicon-based Sentiment Analysis algorithm (Fig. 2). The lexicon used was hand-built from the study of 250 news items. It consists of a structured dictionary of linguistic expressions that can be enhanced with domain extensions. As shall be seen, those extensions contain lexical information on current issues (politics, sports, law, etc.). In the adaptation stage, generic knowledge from the initial lexicon is extended in order to adapt it to the specific domain. This is performed by adding domain-dependent terms and entities. A filter is then applied to the news with a view to excluding all the inappropriate comments so that they will not influence the analysis. Both knowledge in the lexicon and user comments are preprocessed, using various Natural Language Processing (NLP) techniques. Then, an Opinion Focus Detection algorithm is applied to the body of the news and on the comments from users, in order to discover the main discussion topics (opinion focuses). The Sentiment Analysis Module analyzes each sentence in the comments to create a set of tuples that abstractly represents users' opinions by assigning sentiment expressions to opinion focuses. Finally, these tuples are used by the Mining module, which computes the overall sentiment as well as each focus sentiment on the basis of polarity and strength.

In the following sections, each of these stages is discussed in detail. A description of the structure of our lexicon is also provided as well as of the representation mechanisms adopted for storing the linguistic expressions.

4. Linguistic knowledge: lexicon

As mentioned earlier, many of the proposals addressing the problem of Sentiment Analysis use dictionaries or lexicons on which the sentiment expressed in a set of words is mapped. These words, known as **opinion words**, are used with a shallow parsing perspective in order to capture the sentiment of users' sentences. Our lexicon also stores a set of objective expressions (terms that do not express any opinion), which are used to detect the focus references in the comments. The main differences from other lexicons, such as SentiWordNet, is that linguistic expressions are stored as well as single words. In addition, the terms in our lexicon have been collected by hand from real comments allowing us to capture the colloquial language that predominates in user language. Since this colloquial language is full of non-standard expressions, we believe that the analysis of news comments would be incomplete if only standard dictionaries were used. Other authors have already proposed lexicons that rely not on standard dictionaries, supporting colloquial language and multi-word expressions (Velikovich, Blair-Goldensohn, Hannan, & McDonald, 2010).

In the following sections, we first discuss the structure and design of our lexicon, and then explain how the sentences are mapped onto the lexicon.

4.1. Lexicon structure

Many authors have already broadly addressed the problem of classifying words to build lexicons incorporating semantic orientation of individual words and contextual valence shifters (Choi & Cardie, 2008; Danescu-Niculescu-Mizil & Lee, 2010). Moreover several algorithms to enhance initial dictionaries have been proposed and tested (Neviarouskaya et al., 2011). In this study we focus on defining a hierarchical model based on relations between entities. The highlight of our lexicon consists of its practical and interpretable extensible framework.

Our lexicon is structured as a taxonomy of objects and features that represents the most recurrent opinion focuses in the news. An *object* may be a person, an entity, a product, etc. The *features* are heritable characteristics of each object. For example, RESISTENCE, SPEED, and STRENGTH are some of the features of the SPORTSPERSON object. For the sake of simplicity, the term focus is used interchangeably to refer to both objects and features. Capital letters are used in the notation. Fig. 3 shows the section in the lexicon related to SPORTSPERSON and SOCCER_PLAYER. Note that the object FORWARD inherits SIGNING and PASSING features from SOCCER_PLAYER, and RESISTANCE and SPEED from SPORTSPERSON. Even though a taxonomy relations model with objects has previously been used in product reviews (Qiu, Liu, Bu, & Chen, 2009), our taxonomy is different in that no part-of relations (component-subcomponent) are defined. Instead, it uses hierarchical relations, such as in Object-Oriented models (class-subclass).

Additionally, the lexicon contains not only generic valuation expressions but also the particular expressions that users could employ to evaluate a specific focus. For example, "jalopy" is a subjective expression with a negative connotation of the implicit object CAR.

Classical techniques on Sentiment Analysis only consider the polarity of opinions (Negative, Neutral or Positive). According to Wilson, Wiebe, and Hwa (2004), the strength of the opinion should also be represented. For this reason, this study considers valuations such as VERY_NEGATIVE, NEGATIVE, NEUTRAL, POSITIVE, and VERY_POSITIVE.

The lexicon is thus divided into three Classes: (i) a set of hierarchically-related Objects (O); (ii), a set of object Features (F); (iii) a set of Valuations (V).

To select the focuses in our lexicon, we collect the most recurrent discussion topics recent news about sports, politics, economics, current events, and entertainment. The main nodes in our lexicon taxonomy (and some of their child nodes) are the following: PERSON (politician, sportsman, artist, etc.), LEGISLATION (norm, law), ENTITY (institution, group, public organisms, etc.), LOCATION (country, state, etc.), HAPPENING (event, accident, crime, etc.), PRODUCT (entertainment, technology, Internet, etc.), and DISCUSSION_TOPIC (immigration, abort, gay marriage, etc.). Further details on lexicon focuses could be consulted in Appendix A (features were omitted

Table 1
Examples of subjective and objective expressions.

Type	Example expressions	Class
Objective	*Product*	Focus PRODUCT
	Costs	Feature PRODUCT.PRICE
Subjective	*Masterpiece*	Valuation VERY_POSITIVE
	Piece of junk	Focus PRODUCT, valuation NEGATIVE
	A bit expensive	Feature PRODUCT.PRICE, valuation NEGATIVE

Table 2
Modifiers and linguistic rules.

Modifier	Examples	Rules
MOD_VERY	*very*, *absolutely*, *extremely*, ...	MOD_VERY POSITIVE → VERY_POSITIVE
		MOD_VERY NEGATIVE → VERY_NEGATIVE
MOD_LESS	*a little bit*, *relatively*, *not quite*, ...	MOD_LESS VERY_POSITIVE → POSITIVE
		MOD_LESS VERY_NEGATIVE → NEGATIVE
MOD_NEG	*not*, *nothing*, *not at all*, ...	MOD_NEG POSITIVE → NEGATIVE
		MOD_NEG NEGATIVE → POSITIVE
		MOD_NEG VERY_POSITIVE → NEUTRAL[a]
		MOD_NEG VERY_NEGATIVE → NEUTRAL
...	...	...

[a] For example *it is not perfect...*

for simplicity). As shall be seen, it is complex to keep such knowledge updated. For this reason, we have designed a generic lexicon, which only contains abstract knowledge, and various extensions, containing modularized and updated knowledge. Consequently, the general lexicon only includes abstract entities, such as BASKETBALL_TEAM, and lexicon extensions include concrete entities, such as LA_LAKERS, referring to the Los Angeles Lakers basketball team. In order to identify new concrete entities and features to update the extensions some semiautomatic approaches such as Yi et al. (2003), or Zhai, Xu, Kang, and Jia (2011) in chinese texts, could be helpful.

4.2. Linguistic expressions

Each class in the lexicon (O, F, or V) has a set of associated linguistic expressions. The same expression can belong to several classes. Consequently, the following types of expressions are considered: **objective expressions** (not associated with valuations) that refer to objects or features, and **subjective expressions** (associated with valuations) that express a sentiment. Table 1 shows examples of these types of expression. Furthermore, the expressions of a node are inherited by its child nodes, so that if *politician* is an expression for the focus POLITICIAN, it is also an expression for its child node PRESIDENT.

As previously mentioned, our lexicon not only maps single words, but also multi-word expressions. In addition, words in these expressions are not necessarily standard language in contrast to algorithms that extend lists of polarized seed words using WordNet.

As discussed in Qiu et al. (2009), maintaining a universal lexicon for all application domains is (practically) impossible. Attempting to exhaustively store all possible expressions that users could employ to name entities would be unrealistic, especially considering the fact that current issues are constantly changing, and new terms are being coined to refer to products, politicians, athletes, etc. Instead of trying to represent all this knowledge in a static container, our intention is to define an extendible and modularized model. To avoid dependence on the current situation, only abstract focuses are defined in the lexicon and particular focuses are defined in the lexicon extensions. Each concrete entity X defined in an extension is accompanied by an "extends Y" directive, indicating that X is a subclass of Y. In this way, the focuses BASKETBALL_PLAYER, and its child node POWER_FORWARD are defined in the general lexicon, and PAU_GASOL, offspring node of POWER_FORWARD, is defined in a basketball extension. Besides inheriting the expressions associated with the parent focus (e.g. *player* or *power forward*), new objective expressions, such as his name, and subjective expressions, such as his nicknames, should also be added. Moreover, these set of expressions could be enhanced semiautomatically by using some techniques such as Lloyd, Mehler, and Skiena (2006) to detect new objective expressions, Neviarouskaya et al. (2011) and Baroni and Vegnaduzzo (2004) to detect new subjective expressions, or Tan and Wang (2011) to adapt the strength and polarity of subjective expressions from one domain to another.

With this model, the task of keeping the knowledge current becomes easier. For example, let us suppose that in politics, the opposing party Y wins the elections, and the political party X then becomes the opposition. In this case, the knowledge could be updated just by modifying the hierarchical relations between objects (Fig. 4).

Contextual valence shifters (e.g. negators, adverb intensifiers, etc.) play a key role in the improvement of the representation of the expressions. Those kind of linguistic modifiers are also supported on this study. Table 2 briefly shows some examples of how linguistic patterns display a larger language. To access a vaster discussion on this mechanism Choi and Cardie (2008) and Danescu-Niculescu-Mizil and Lee (2010) could be consulted.

5. Algorithm of analysis

This section provides a detailed explanation of the analytical method. First, the preparation stage, which involves preprocessing and filtering comments, is described. Then we go onto explain how the lexicon is used to identify focuses and how the Sentiment Analysis Module classifies them. Finally, the mining module and the interpretability of the results are described.

5.1. Preparation of the knowledge

In this section, the preparation of the Knowledge Module is explained. This module includes a filtering stage and a preprocessing stage.

5.1.1. Filtering stage

The purpose of the Filter Module is to identify and automatically dismiss the noisy (offensive, non-related or advertising)

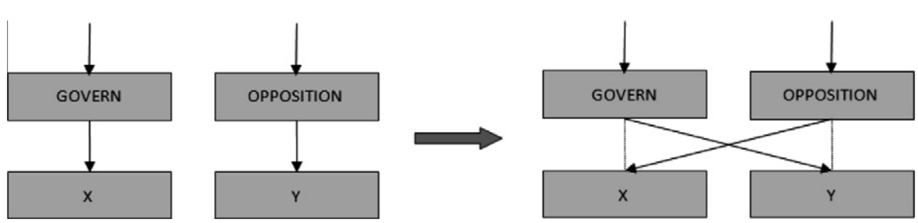

Fig. 4. Example of a political scenario and knowledge maintenance.

Table 3
Example 2: Detecting local context of a comment.

Trace		$\mathcal{C}_l$
Comment: "… What's wrong with this man! Always the same movie. This nation needs urgently a change of president, someone with grater leadership capacity! Spain needs that its politicians improve the current situation and the opposition party does nothing!"		Ø
2.a Initial focus detection	*man* → PERSON *movie* → MOVIE *nation* → COUNTRY *president* → PRESIDENT *politicians* → POLITICIAN *Spain* → SPAIN *opposition party* → OPPOSITION	{PERSON, MOVIE, COUNTRY, PRESIDENT, POLITICIAN, SPAIN, OPPOSITION}
2.b Disambiguation analysis	**root/person** **root**/PRODUCT/ENTERTAINMENT/MOVIE **root/location/country** **root/person**/POLITICIAN/PRESIDENT **root/person**/POLITICIAN **root/location/country**/SPAIN **root/person**/POLITICIAN/OPPOSITION	{MOVIE, PRESIDENT, POLITICIAN, SPAIN, OPPOSITION}
2.c Frequency analysis	**f (movie)=Very low** f (PRESIDENT)=High f (POLITICIAN)=High f (SPAIN)=Medium **f (OPPOSITION)=Low**	{PRESIDENT, POLITICIAN, SPAIN}
2.d Disambiguation analysis II	**root/person/politician** /PRESIDENT **root/person/politician** ROOT/LOCATION/COUNTRY/SPAIN	{PRESIDENT, SPAIN}
2.e Feature discovery	*leadership capacity* → PRESIDENT.LEADERSHIP	{PRESIDENT, PRESIDENT.LEADERSHIP, SPAIN}

comments so that they will not influence the analysis. The following types of comments were regarded as inappropriate, and thus were filtered out:

In colloquial language, swear words are very common. This type of comment should be revised carefully since discarding them all would lead to an excessive loss of information and be detrimental to the analysis. For this reason, only comments containing potentially offensive words that quote other user's comment are deleted, because it is very likely that they are expressing an opinion about another user instead of about the news focuses.

Users generally insert links in their comments. Often, those comments have advertising purposes. Therefore, it is likely that they will be accompanied by positive terms that should not influence the analysis (*Sign in now at < advertising url > where you will find the best prices!*). The elimination of those advertising comments is a heuristic used in order to eliminate opinions on irrelevant topics.

Certain media allows users to report inappropriate behavior from other users. This filter automatically eliminates these comments.

5.1.2. Preprocessing stage

In the preprocessing stage, various NLP techniques are applied. First we use the PoS tagger (Manning & Schutze, 1999) to improve the performance of the Stemmer. By stemming the comments and linguistic expressions stored in the lexicon, generality is gained. Finally, the Splitter is applied to separate the sentences of the comments. The GPL Library FreeLing 2.2[2] was used to implement the preprocessing.

5.2. Opinion focus detection module

In the Opinion Focus Detection Module, the lexicon detects which focuses are the object of valuations. To illustrate this, an example of a real comment on politics is now considered (Example 2 in Table 3). To simplify tracking, this comment has not been preprocessed.

5.2.1. Interpretation context

Context is defined here as the set of focuses that disambiguate the interpretation of subjective expressions. For instance, "big" has a negative connotation if it is used to evaluate the feature SIZE of the object MOBILE_PHONE, but it has a positive connotation if it is used to express an opinion about the feature SIZE of the object TV. For this reason, delimiting the context (MOBILE_PHONE or TV in this example) is crucial. The context could be analyzed from a comment-level (local context $\mathcal{C}_l$) or document-level (global context $\mathcal{C}_g$).

A focus is considered to be implicit if it does not explicitly appear in an opinion-bearing sentence. However, as previously mentioned, in our model, subjective expressions could also be attached to objects and features. There are cases when more than one focus is attached to the subjective expression. The use of information in the local context to disambiguate the focus is possible in most cases.

Firstly, the objects in the lexicon containing some expression that appears in the document (body or comments) are marked as initial candidates (Example 2.a) in order to initialize the interpretation context. It is very likely that this initial set may be too broad because of the ambiguity of certain expressions. To reduce the candidate set, two heuristics are applied: **disambiguation analysis** and **frequency analysis**.

5.2.2. Disambiguation analysis

The disambiguation analysis uses the implicit knowledge contained in the hierarchical structure of the lexicon to tackle with the focus co-reference. For example, since only PERSON and POLITICIAN objects were detected as candidates, this analysis discards PERSON from the candidate set, because it is an antecessor (a co-reference) of POLITICIAN in the lexicon hierarchy, which is more specific. Note that POLITICIAN is kept because it could be a co-reference of PRESIDENT or OPPOSITION (Example 2.b).

More formally, the disambiguation analysis discards all the focuses f in the local context $\mathcal{C}_l$ that are in the set defined by $\{f \in \mathcal{C}_l \mid \exists f' \in (desc(f) \cap \mathcal{C}_l) : (desc(f) \cap \mathcal{C}_l - \{f'\}) \subseteq desc(f')\}$, where $desc(x)$ denotes the set of all the descending focuses of x in the lexicon hierarchy.

[2] http://nlp.lsi.upc.edu/freeling/.

5.2.3. Frequency analysis

The frequency analysis, discards low-referenced focuses in the entire set of comments. This avoids noisy comments that make it difficult to interpret the results (Example 2.c).

Unfortunately, this process is often affected by situations where the ambiguity is more complex and the previously mentioned analyses are not sufficient. For example, if POLITICIAN, PRESIDENT, and MEMBER_OPPOSITION are in the candidate set, POLITICIAN cannot be discarded because both PRESIDENT and MEMBER_OPPOSITION focuses are descendants of POLITICIAN in the lexicon. Accordingly, an expression referring to the POLITICIAN focus would be ambiguous since it could refer to either focus. However, as some of the focuses in the local context could have been deleted after applying the frequency analysis, the disambiguation analysis could possibly discard more focuses (Example 2.d).

Since both heuristics eliminate focuses, they perform simultaneously until there are no more changes to be made in $\mathcal{C}_l$. After that, features are searched by identifying objective expressions associated with any features of the object focus in $\mathcal{C}_l$ (Example 2.e). Finally, the global context $\mathcal{C}_g$ is created as the union of each local context.

As shall be seen in the next section, focuses that achieve a low reliability analysis are also deleted from $\mathcal{C}_g$ in the mining stage, and their presence in the results summary is thus avoided.

5.3. Sentiment Analysis Module

In order to analyze the overall sentiment of the comments and also assign sentiment expressions to opinion focuses, we designed a Sentiment Analysis algorithm that is composed of three stages: (i) Expression Labeling; (ii) Tuples Extraction; (iii) Tuples Clustering and Filtering. Much of the current research on Sentiment Analysis focuses on document-level analysis and sentence-level analysis. However, since comments represent our basic information unit, we also perform an intermediate comment-level Sentiment Analysis.

To facilitate explanation, each stage is applied to a real comment composed of three sentences (Example 3), after the following global context has been previously detected: $\mathcal{C}_g$:={LA_LAKERS, NY_NICKS, TEAM, L_WALTON, L_ODOM, FORWARD, BASKETBALL_PLAYER, SPORTSMAN, DRIBBLING, PASSING}.

```
(s1)  Things are breaking well for The Lakers
(s2)  It's a formidable team
(s3)  Odom is a great forward, but I think that in
      dribbling and passing, Walton is the best
```

Example 3. Comment extracted from sport news.

5.3.1. Labeling expression stage

In the Labeling Expression Stage, every expression in each comment that could refer to a focus of the global context is automatically identified and labeled. Taking the context and the valuation class V as a starting point, all the linguistic expressions associated with them in the lexicon are consulted. It should be underlined that each expression of a node is inherited by their offspring nodes. These expressions are used to detect and label references to the focuses in the comments. In a similar way, the valuations are marked. After labeling the sample comment (Example 3.b), some expressions are marked with more than one label, such as *forward*, which can refer to the focuses FORWARD, L_WALTON or LAMAR_ODOM.

(s1) [Things are breaking well]$_{POSITIVE}$ for [The Lakers]$_{LA\_LAKERS}$.
(s2) It's a [formidable]$_{VERY\_POSITIVE}$ [team]$_{TEAM\text{ or }LA\_LAKERS\text{ or }NY\_KNICKS}$.
(s3) [Odom]$_{L\_ODOM}$ is a [great]$_{POSITIVE}$ [forward]$_{FORWARD\text{ or }L\_ODOM\text{ or }L\_WALTON}$, but I think that in [dribbling]$_{DRIBBLING}$ and [passing]$_{PASSING}$, [Walton]$_{L\_WALTON}$ is [the best]$_{VERY\_POSITIVE}$.

Example 3.b. Sentences labeling.

5.3.2. Tuples extraction stage

In this stage, all of the analysis tuples for each sentence are computed. Such tuples atomically capture each user's opinion and are used for further mining. Before describing the process, tuples of analysis must be defined more formally.

An analysis tuple T is a tern $\langle o,f,v\rangle$ where $o \in O$ is an object satisfying $(o \in \mathcal{C}_g) \vee o = \emptyset$; $f \in F$ is a feature satisfying $(f \in (features(o) \cap \mathcal{C}_g)) \vee f = \emptyset$, and finally, $v \in V$ is a valuation that can also be null. Although other authors also consider parameters t (date of the opinion) and h (opinion holder, also called opinion sources) (Wiebe et al., 2005), those parameters are irrelevant to our problem because users often report their comments on a date close to the news and tend to remain anonymous.

The algorithm is initialized by replacing each labeled expression with a tuple. When more than one label is possible, the most general one is selected (i.e. the one that corresponds to the first common predecessor). The distance that separates each pair of consecutive tuples, which is measured as the number of non-labeled words between them, is recorded. Example 3.c shows the third sentence of the example comment after initialization.

$$[\langle \text{L\_odom},\emptyset,\emptyset\rangle\ \mathbf{2}\ \langle\emptyset,\emptyset,\text{POSITIVE}\rangle\ \mathbf{0}\ \langle\text{FORWARD},\emptyset,\emptyset\rangle\ \mathbf{5}\ \langle\text{BASKETBALL\_}\mathit{player},\ \text{DRIBBLING},\emptyset\rangle\ \mathbf{1}\ \langle\text{BASKETBALL\_}\mathit{player},\ \text{PASSING},\emptyset\rangle\ \mathbf{0}\ \langle\text{L\_walton},\emptyset,\emptyset\rangle\ \mathbf{1}\ \langle\emptyset,\emptyset,\text{VERY\_}\mathit{positive}\rangle]$$

Example 3.c. Initialization of the tuples array and distances for sentence s3.

The algorithm joins the tuples by proximity. It follows the heuristic that one valuation is likely by referring to the nearest focus (Table 4, Algorithm 1). The pseudo-distance -1 is used to indicate that two tuples were already joined. The process ends when all the distances between tuples are -1.

Where t_i is the ith tuple and $n_i = distance(t_i, t_{i+1})$ in the list of tuples L. In lines 2–3, the simple joining of focuses and valuations is performed. This union follows the heuristic that a valuation followed by a focus is likely to be a direct valuation of this focus. In lines 4–5, two close tuples are joined by verifying their hierarchical relationship. If they are hierarchically related, the most specific focus replaces the other. Where $t_i = \langle o_i, f_i, v_i\rangle$ and

Table 4
Algorithm 1: Tuples extraction.

0:	**procedure** Extraction ($L = [t_0, n_0, t_1, n_1, \ldots, t_{k-1}, n_{k-1}, t_k]$)
1:	**do**:
2:	**foreach** $\{t_i = \langle\langle\emptyset,\emptyset,v_i\rangle, n_i = 0, t_{i+1} = \langle f_{i+1}, p_{i+1}, \emptyset\rangle\}$
3:	$L \leftarrow t_0, \ldots, n_{i-1}, \langle f_{i+1}, p_{i+1}, v_i\rangle, n_{i+1}, \ldots, t_k$
4:	**foreach** $\{n_i = 0\}$
5:	$L \leftarrow t_0, \ldots, n_{i-1}, U(t_i, n_i, t_{i+1}), n_{i+1}, \ldots, t_k$
6:	**while** $\exists([\ldots \langle f_i, p_i, v_i\rangle, -1, \langle f_{i+1}, p_{i+1}, v_{i+1}\rangle \ldots])$ *AND* $(v_i = \emptyset$ *XOR* $v_{i+1} = \emptyset)$
7:	if $(v_i = \emptyset) v_i \leftarrow v_{i+1}$
8:	else $v_{i+1} \leftarrow v_i$
9:	$n_{min} := min_i\{n_i > 0\}$
10:	**foreach**$n_i \neq -1$
11:	$n_i \leftarrow n_i - n_{min}$
12:	**while** $\{\exists n_i > -1\}$

Table 5
Example 3.d Assignation of sentiment to focuses trace in s3.

Trace	Explanation
[⟨L_ODOM, ø, ø⟩ 2 ⟨**forward**, ø, **positive**⟩ 5 ⟨BASKETBALL_PLAYER, DRIBBLING, ø⟩ 1 ⟨BASKETBALL_PLAYER, PASSING, ø⟩ 0 ⟨L_WALTON, ø, ø⟩ 1 ⟨ø, ø, VERY_POSITIVE⟩]	Simple U.
[⟨L_ODOM, ø, ø⟩ 2 ⟨FORWARD, ø, POSITIVE⟩ 5 ⟨BASKETBALL_PLAYER, DRIBBLING, ø⟩ 1 ⟨**l_walton**, PASSING, ø⟩ **−1** ⟨**l_walton**, ø, ø⟩ 1 ⟨ø, ø, VERY_POSITIVE⟩]	U
[⟨L_ODOM, ø, ø⟩ **1** ⟨FORWARD, ø, POSITIVE⟩ **4** ⟨BASKETBALL_PLAYER, DRIBBLING, ø⟩ **0** ⟨L_WALTON, PASSING, ø⟩ −1 ⟨L_WALTON, ø, ø⟩ **0** ⟨ø, ø, VERY_POSITIVE⟩]	Bring tuples 1 unit closer
[⟨L_ODOM, ø, ø⟩ 1 ⟨FORWARD, ø, POSITIVE⟩ 4 ⟨**l_walton**, DRIBBLING, ø⟩ **−1** ⟨**l_walton**, PASSING, ø⟩ −1 ⟨**l_walton**, ø, **very_positive**⟩]	Simple U. and U
[⟨L_ODOM, ø, ø⟩ 1 ⟨FORWARD, ø, POSITIVE⟩ 4 ⟨**l_walton**, DRIBBLING, VERY_POSITIVE⟩ **−1** ⟨**l_walton**, PASSING, VERY_POSITIVE⟩ **−1** ⟨**l_walton**, ø, **very_positive**⟩]	Valuation Propagation
...	...
[⟨L_ODOM, ø, POSITIVE⟩ −1 ⟨L_WALTON, DRIBBLING, VERY_POSITIVE⟩ −1 ⟨L_WALTON, PASSING, VERY_POSITIVE⟩ −1 ⟨L_WALTON, ø, VERY_POSITIVE⟩]	Final

$t_{i+1} = \langle o_{i+1}, f_{i+1}, v_{i+1}\rangle$ are the input tuples, rule $U(t_i, n_i, t_{i+1})$ modifies $t_i \leftarrow \langle U_o(o_i, o_{i+1}) f_i,\ U_v(v_i, v_{i+1})\rangle$ and $t_{i+1} \leftarrow \langle U_o(o_{i+1}, o_i), f_{i+1}, U_v(v_{i+1}, v_i)\rangle$ returning the sublist $[t_i, -1, t_{i+1}]$ where U_o represents the union of objects Eq. (1) and U_v, the union of valuations Eq. (2). After applying rule U, duplicate tuples are eliminated. In lines 6–8, valuations are propagated to the left and right through the already joined tuples in order to avoid the influence of the order in the result. Finally, the next pair of the nearest tuples is searched in line 9, and their distance is then subtracted from all other distances in lines 10–11. Example 3.d shows the trace of the algorithm applied to the previous sample array of tuples. In this way, the opinion expression *great* is assigned to Odom, and *the best* is assigned to Walton, also evaluating the features PASSING and DRIBBLING.

$$U_o(o_x, o_y) = \begin{cases} o_y & \text{iff } o_y \in desc(o_x) \\ o_x & \text{in other case} \end{cases} \tag{1}$$

$$U_v(v_x, v_y) = \begin{cases} v_y & \text{iff } v_x = ø \\ v_x & \text{in other case} \end{cases} \tag{2}$$

5.3.3. Tuples clustering and filtering stage

In this stage, all the focuses extracted for each sentence of a comment, are clustered, and a single set of tuples that abstractly represents user opinion is created.

In this process, the disambiguation analysis is applied again, but only considering the focus involved in the comment, namely, $\mathcal{C}_l$ instead of the entire context $\mathcal{C}_g$. For example, considering the tuples and extracted from s1 and s2, the result of this analysis replaces TEAM with LA_LAKERS due to the fact that it is not involved in the comment even though NY_NICKS is also in the global context. Finally, the result of applying the Sentiment Analysis to the sample is {⟨LA_lakers, ø, POSITIVE, ⟨LA_lakers, ø, VERY_positive⟩, ⟨L_odom, ø, POSITIVE, ⟨L.WALTON, DRIBBLING, VERY.POSITIVE, ⟨L.WALTON, PASSING, VERY_*positive*, ⟨L_WALTON, ø, VERY_POSITIVE⟩}

5.4. Sentiment mining module

Once the set of tuples for each comment has been obtained, any mining technique is feasible. We compute the average sentiment for each focus involved in the comments. The general sentiment in the entire news is measured as an average of the specific opinions reported for each particular focus in the comments. To compute mean values, the following weightings are assigned to valuation labels: VERY_NEGATIVE:=-10, NEGATIVE:=-5, NEUTRAL:=0, POSITIVE:=5 y VERY_POSITIVE:=10.

Finally, a sentiment report is obtained (Fig. 1). This report contains the overall sentiment, and the focus sentiment for each focus in the global context. Each feature sentiment is accompanied by a reliability measurement, obtained as the percentage of comments that have provided information related to this feature. Furthermore, not only the average sentiment is reported, but also the partial sentiments given to each possible valuation. This allows the study of polarity distributions in the users' opinion. As previously mentioned, in this stage, the latter focus filter is applied, and focuses whose reliability measurement is lower than 5% are discarded.

6. Evaluation performance

This section presents the computational results obtained in the experiments that were performed to evaluate the LCN-SA system.

6.1. System evaluation

Our system was evaluated by means of a broad test set of comments about current news items. The resulting reports were manually validated by 20 volunteers. They were requested to read the comments and then check the consistency of the reports obtained. It was found that in 87.6% of all cases, the results were fully consistent. In 7.4% of the cases the analysis was incomplete, and in the remaining 5%, the results were incorrect. More detailed information of this evaluation grouped by categories is shown in Table 5. However, those results are not conclusive enough for the following reasons: (i) agreeing with the volunteers feedback, it is hard and costly for human to determine the analysis of general-domain sets of comments; (ii) since we intend to validate a whole system, those results remind insufficient.

Since our system includes clearly differentiated stages, we propose a module comparison against some of the most representative SENTIMENT ANALYSIS algorithms rather than a direct comparison by means of a black-box evaluation. Table 6 summarizes the capabilities of our system and some of the most representative Sentiment Analysis approaches: Machine-learning (ML) Approaches (Pang et al., 2002), PMI-IR (Turney, 2002), Sentiment Analyzer (SA) (Yi et al., 2003), FBS (Hu & Liu, 2004a), and SO-CAL (Taboada et al., 2011).

Both our system and SO-CAL are lexicon-based approaches that require a considerable human effort to build the lexicon. This effort is essentially equivalent to providing annotated training resources in ML methods. According to Taboada et al. (2011), lexicon-based methods perform robustly in cross-domain scenarios, and can be easily enhanced with external sources of knowledge. There are methods for automatically adapting polarity lexicons to new domains (Qiu et al., 2009). In contrast, ML methods require more effort to annotate new data resources.

6.2. Module comparison

Our method is mainly composed of two modules (the Focus Detection Module and the Sentiment Analysis Module). In order to objectively validate our system, we designed a set of experiments that allowed us to evaluate each module separately (crystal-box evaluation) for news comments datasets. Additionally, the overall sentiment of the entire news corpus is also given so that it can be compared to the results of conventional Sentiment Analysis methods which only calculate overall polarity. Unfortunately, the properly annotated datasets are not available. Furthermore, a

Table 6
Experts Validation Summary.

	Sports	Politics	Economy	Society	Entertainment
Correct (%)	87	88	93	84	86
Incomplete (%)	8	9	4	7	9
Incorrect (%)	5	3	3	9	5

Table 7
Comparison of the capabilities of the main related algorithms.

Capabilities	LCN-SA	ML	PMI-IR	SA	FBS	SO-CAL
Overall sentiment	**Y**	Y	Y	Y	Y	Y
Feature mining	**Y**	N	N	Y[a]	Y	Y
Topic feature discovery	**Y**	N	N	Y	Y	N
Strength measure	**Y**	N	Y	N	N	Y
Multidomain	**Y**	N	N	N	N	Y
Expert supervision	**Y**	Y	N	N	Y[b]	Y

[a] Although the algorithm is able to perform Feature Mining, the lexicon is not available.
[b] A training dataset is required in CBA (Liu et al., 1998) exclusively in the Topic features discovery module. It is not necessary for the other modules.

comparison against SO-CAL was not even considered since this method and ours are complementary. Indeed, we plan to mix up our lexicon with SO-CAL methodology in further research. Thus, we selected the following comparison algorithms and experiments for each module:

- **Focus Detection Module:**
  - **The Sentiment Analyzer** consists of three main modules: (i) candidate feature module; (ii) detection module; (iii) feature selection and Sentiment Analysis Module. The last stage uses a lexicon that is unavailable. For this reason, only the candidate feature selection and the feature selection modules are considered in the comparison. The authors explain three methods of extracting noun phrases following different heuristic patterns (BNP base noun phrases, dBNP definite base noun phrases, and bBNP beginning definite base noun phrases). We considered the bBNP because it was demonstrated that this heuristic outperformed the others.
- **Sentiment Analysis Module applied to focuses (Focus SA):**
  - **FBS (Feature-Based Summarization)** uses the association miner CBA algorithm to obtain candidate features. The CBA classifier is not considered here because it requires a labeled training dataset of news. However, to obtain the sentiment in each feature, it applies an algorithm that exploits synonymy and antonymy relations in WordNet to expand the initial set of polarized words (seed adjectives). Then, the polarity of each feature is calculated.
- **Sentiment Analysis Module applied to the entire document (Overall SA):**
  - **PMI-IR (Pointwise Mutual Information – Information Retrieval)** is one of the most widely used algorithms in the literature. It is used to classify the polarity of the entire document.
  - **FBS** also classifies the entire document as positive, neutral, or negative.

6.2.1. The data

The experiments were performed on a set of 500 current news items randomly selected from various news media. These news articles were manually labeled by 10 student volunteers who were unaware of the structure or the content in the lexicon. For each news item, the following information was requested to the students to reflect user opinions:

Table 8
News distribution by kind and average number of user comments.

	Sports	Politics	Economy	Society	Entertainment
Percentage (%)	27	18	13	25	17
Comments average	301.5	177.1	146.1	158.5	142.7

- Classification of general user opinions in {VERY_NEGATIVE, NEGATIVE, NEUTRAL, POSITIVE, VERY_POSITIVE}.
- Set of the main discussion topics (focuses) in each news comments.
- Discussion topics (focuses) involved in each comment and the valuation that summarizes the user's opinion of each one.

The documents concerning the creation of the lexicon and the documents involved in the experimental evaluation were randomly extracted from the news media, 20 Minutos[3] and Mail Online[4] from 05/05/2010 to 06/07/2011. The news items chosen belong to a wide variety of categories (sports, politics, economics, society and culture) and each contains at least 50 comments (Table 7). It should be underlined that none of these news items intervened in the creation of the lexicon. Our lexicon built by means of the manual analysis of 2442 comments in 250 news, contains 380 objects (including 182 abstract focuses and 198 concrete focuses) and 128 features, with a total of 4762 linguistic expressions (1516 objective expressions and 3246 subjective expressions) (see Table 8).

6.2.2. Performance of the experiments

Using the labeled news as prototypes, the Focus Detection Module and the Sentiment Analysis Module were evaluated using the F-measure metric (Eq. (5)). This metric is the weighted harmonic mean of precision (Eq. (3)) and recall (Eq. (4)), which are the most common metrics used in the literature. Precision is the fraction of the focuses detected that are considered relevant. Recall is the fraction of the focuses detected that are relevant to the labeled focuses.

$$Precision = \frac{\#(\{labeled\ focuses\} \cap \{retrieved\ focuses\})}{\#\{retrieved\ focuses\}} \tag{3}$$

$$Recall = \frac{\#(\{labeled\ focuses\} \cap \{retrieved\ focuses\})}{\#\{labeled\ focuses\}} \tag{4}$$

$$F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{5}$$

To evaluate the Overall Sentiment Analysis, the Accuracy metric is used, which considers the polarity of the prototype and the polarity calculated (Eq. (6)). The accuracy of a measurement system is the degree of closeness of the measurements of a quantity to its actual (true) value.

$$Acc = \frac{\#\{succesfully\ labeled\ news\}}{\#\{news\}} \tag{6}$$

6.2.3. Results of the study of the news datasets

This section discusses the results obtained for each module and the comparison with the selected algorithms. Table 9 shows the results obtained by our Focus Detection Module as compared to the Sentiment Analyzer. Table 10 shows the results of our Sentiment Analysis Module in the Overall SA experiment in comparison to the PMI and FBS. Table 11 shows the results of the Sentiment Analysis Module in the Focus SA as compared to the FBS. Finally, Fig. 5 provides a summary of the comparison results obtained for each experiment.

In the Focus Detection experiment, both the Sentiment Analyzer

[3] www.20minutos.es.
[4] http://www.dailymail.co.uk.

Table 9
Focus detection results.

	News																	
	Sports			Politics			Economy			Society			Entert.			Total		
	P	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1
LCN-SA	.89	.78	.84	.83	.82	.82	.94	.82	.88	.75	.71	.73	.69	.62	.65	.82	.75	.78
S.Analy.	.79	.62	.84	.70	.62	.66	.82	.73	.77	.68	.66	.67	.67	.67	.67	.73	.65	.69

Table 10
Overall Sentiment Analysis results.

	News					
	Sports	Politics	Economy	Society	Entert.	Total
LCN-SA	.68	.98	.98	.96	.94	.89
PMI	.43	.99	.98	.92	.88	.81
FBS	.54	.51	.77	.76	.64	.64

and our system achieved good results. The Sentiment Analyzer has the advantage of operating without any external knowledge. However, in view of the results, our structured lexicon, which is used to detect discussion topics, appeared to perform better. With respect to the overall sentiment experiment, results indicate that calculating the overall sentiment based on the particular sentiment of each discussion topic was a better approach to the news analysis problem than the performance of a global calculation as is the case of the PMI.

However, the fact that we were obliged to select comparison algorithms that were specifically designed for product reviews means that the results should be interpreted accordingly. Although the starting point of Sentiment Analysis is common to both approaches, the news to which the analysis was applied has a number of difficulties for which our algorithm is particularly robust. The multi-domain scenario motivates context detection and disambiguation processes as well as the use of colloquial language in news comments, all of which are examples of such difficulties. In this regard, we calculated that only 40.75% of the sentences in our news data set contained adjectives belonging to standard language. The same is also true of various other scenarios, such as debate forums, social networks, and blogs. This proliferation of colloquial language means that in such scenarios, the FBS algorithm is problematic to apply since it would be forced to work without sufficient information. However, this difficulty is specific of this kind of user-generated content. It goes without saying that the validity of the FBS in product reviews has been tested and confirmed (Hu & Liu, 2004a).

It should be underlined that the results shown were obtained after applying the Filtering Stage. We have computed that 7.561% of the comments were discarded. Of these, 57.248% correspond to comments containing swear words and 42.752% correspond to comments containing URLs. Since most of the discarded comments involved swear words (with negative sentiment), the influence of discarding them usually reminds in a positive variation of the overall sentiment score. Indeed, we have computed a positive variation of around 2.6% in our experiments.

Table 11
Focus Sentiment Analysis results.

	News																	
	Sports			Politics			Economy			Society			Entert.			Total		
	P	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1
LCN-SA	.93	.63	.75	.80	.61	.70	.91	.72	.80	.80	.62	.70	.85	.60	.71	.86	.63	.72
FBS	.61	.33	.75	.59	.38	.46	.63	.42	.50	.56	.35	.43	.76	.37	.50	.62	.36	.45

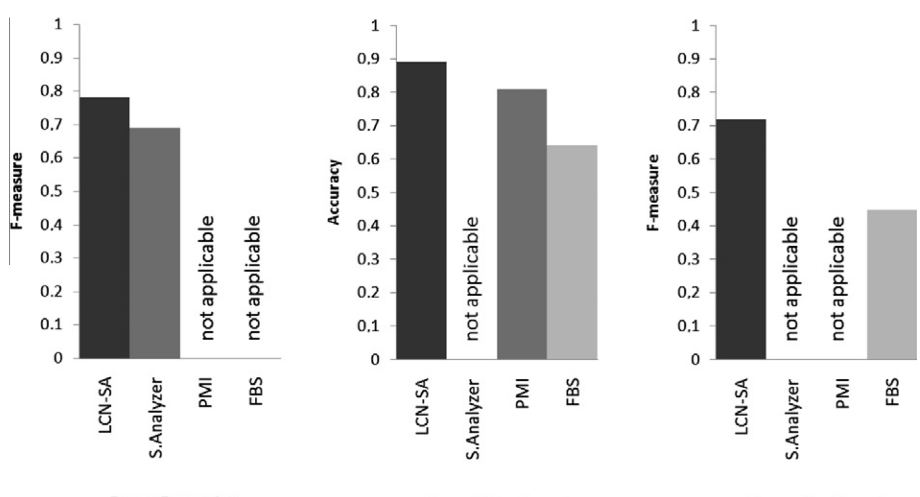

Fig. 5. Result summary comparative.

7. Conclusions and future work

This article has proposed a complete lexicon-based system to the sentiment analysis of user comments on current news items. This type of analysis has several particularities, such as its multi-focus scenario or the use of colloquial language. Our system includes a Focus Detection Module that is able to identify the main discussion topics. It also contains a Sentiment Analysis Module which is able to analyze the strength and sentiment of the entire news article as well as of each of its focuses. Both modules use a lexicon specifically built to analyze news comments.

The system has obtained promising results in experimental validation. Given that our system includes various modules, we designed specific experiments to compare each module with some of the most relevant algorithms in Sentiment Analysis. The results obtained in these experiments showed that our approach performed better in scenarios where colloquial language predominates.

The system could be easily applied to other domains. Although the lexicon that we used is specifically designed to deal with news comments, the knowledge in our system is highly modularized. By adapting the knowledge, the system could carry out Sentiment Analysis in other domains.

One of the aims of this research was also to find a way to adapt knowledge. Our lexicon consists of a set of hierarchically-structured objects and features. Two types of entities were taken into account: (i) abstract entities that remain invariable in the lexicon; (ii) concrete entities that are included in the extensions. Undoubtedly, the task of designing and building a structured lexicon takes much more effort than automatically expanding sets of opinion words through WordNet relations. However, our lexicon has certain important advantages, as shown in the results: (1) it not only contains single standard language words but also colloquial expressions that are of crucial importance in the analysis of real comments; (2) its hierarchical structure contains implicit information that can be exploited in order to improve the performance of the focus detection through disambiguation analysis; (3) the lexicon extensions allow current dependent knowledge to be modularized. Furthermore, they can be easily maintained just by adding or modifying some of their objects, features, or hierarchical relations.

Although results are promising, there is much work still to be done. Exploiting semantic relations through WordNet has proven to be a successful heuristic for automatically expanding opinion word sets. In this line, in future work, we plan to study the automatic expansion of subjective expressions of our lexicon, or even the automatic detection of new entities through web mining.

There is no doubt that Shallow Language approaches lead to lighter analyses than NLP or other techniques, and their effectiveness has been proven in several fields. However, these techniques are not able to detect shades of meaning, such as irony or sarcasm. For this reason, it is difficult to accurately calculate sentiment in such cases. In further research, our aim is to study these shades of meaning, and thus improve our Sentiment Analysis Module.

Due to their interest and potential practical application, product reviews have received a great deal of attention. Since our knowledge model can be easily applied to other domains, we plan to modify the set of focuses in the lexicon and test our system in this area. Since comparative opinions and temporal analysis are crucial in product reviews, they should certainly be addressed.

Another goal for future research is the definition of a more effective context model in order to improve the disambiguation analysis and also to address ellipsis and anaphora. To achieve this, we propose the study of the order of the comments (comments history) to define a dynamic context of the discussion topics, which can serve to improve sentiment analysis in cases where there is ellipsis or anaphora.

Acknowledgments

The authors thank the *Ministerio de Educación y Ciencia* and *Junta de Andalucía* that supported this article with its Projects: TIN2007-60199, TIC2009-5011 and TIN2007-67984

Appendix A. Lexicon hierarchy

```
<LexiconHierarchy>
 <Focus="NEW">
  <Focus="PERSON">
   <Focus="POLITICIAN">
    <Focus="CABINET_MEMBER">
    <Focus="PRESIDENT">
    <Focus="OPPOSITION">
     <Focus="OPPOSITION_LEADER">
    <Focus="FORMER_PRESIDENT">
   <Focus="SPORTMAN">
    <Focus="SOCCER_PLAYER">
     <Focus="GOAL_KEEPER">
     <Focus="FORWARD">
     <Focus="DEFENSE">
    <Focus="MIDFIELD">
    <Focus="TENNIS_PLAYER">
    <Focus="BASKETBALL_PLAYER">
    <Focus="FOOTBALL_PLAYER">
    <Focus="HANDBALL_PLAYER">
    <Focus="GOLFER">
    <Focus="DRIVER">
     <Focus="F1">
     <Focus="MOTO">
    <Focus="CYCLIST">
    <Focus="ATHLETE">
    <Focus="REFEREE">
    <Focus="COACH">
   <Focus="BULLFIGHTER">
   <Focus="ACTOR">
   <Focus="DIRECTOR">
   <Focus="ARTIST">
    <Focus="SINGER">
    <Focus="GUITARRIST">
    <Focus="DRUMMER">
    <Focus="BASSIST">
    <Focus="DANCER">
    <Focus="PAINTER">
    <Focus="COMPOSER">
    <Focus="MODEL">
   <Focus="CELEBRITY">
   <Focus="EMPLOYEE">
    <Focus="STATE_OFFICIAL">
   <Focus="TEACHER">
   <Focus="BUSSINESSMAN">
    <Focus="SPORTS_TEAM_OWNER">
   <Focus="TEENAGER">
   <Focus="VICTIM">
   <Focus="CRIMINAL">
    <Focus="TERRORIST">
     <Focus="TERRORIST_LEADER">
   <Focus="THIEF">
   <Focus="KILLER">
   <Focus="AGRESSOR">
 <Focus="LEGISLATION">
  <Focus="LAW">
```

(*continued on next page*)

```
      <Focus="ESTATUTE_OF_AUTONOMY">
    <Focus="REGULATION">
  <Focus="ENTITY">
    <Focus="ORGANIZATION">
      <Focus="POLITICAL_PARTY">
        <Focus="GOVERNMENT">
        <Focus="OPPOSITION">
      <Focus="SINDICATE">
      <Focus="COMPANY">
        <Focus="TV_CHANNEL">
        <Focus="RECORD_LABEL">
      <Focus="STOCK_MARKET">
      <Focus="TERRORIST_ORGANIZATION">
    <Focus="PUBLIC_ORGANISM">
      <Focus="SECURITY_FORCES">
        <Focus="POLICE">
        <Focus="CIVIL_GUARD">
    <Focus="INSTITUTION">
      <Focus="EXECUTIVE_BRANCH">
      <Focus="LEGISLATIVE_BRANCH">
        <Focus="MINISTRY">
      <Focus="UNIVERSITY">
      <Focus="CHURCH_INSTITUTE">
    <Focus="TEAM">
      <Focus="CLUB">
      <Focus="F1_TEAM">
    <Focus="GROUP">
      <Focus="RESEARCH_GROUP">
      <Focus="BAND">
  <Focus="LOCATION">
    <Focus="CONTINENT">
      <Focus="EUROPE">
      <Focus="AFRICA">
      <Focus="ASIA">
      <Focus="AMERICA">
      <Focus="OCEANIA">
      <Focus="ANTARCTICA">
    <Focus="COUNTRY">
    <Focus="COMUNITY">
    <Focus="PROVINCE">
    <Focus="CITY">
    <Focus="TOWN">
  <Focus="HAPPENING">
    <Focus="INCIDENT">
      <Focus="NATURAL PHENOMENOM">
        <Focus="GLOBAL_WARMING">
        <Focus="NATURAL_DISASTER">
      <Focus="ACCIDENT">
      <Focus="ROBBERY">
      <Focus="GENOCIDE">
      <Focus="TERRORIST_ATTACK">
      <Focus="MURDER">
      <Focus="AGRESSION">
      <Focus="FRAUD">
      <Focus="CORRUPTION">
      <Focus="GENDER_VIOLENCE">
      <Focus="DISCOVERY">
      <Focus="ARREST">
    <Focus="EVENT">
      <Focus="SPORT_EVENT">
        <Focus="MATCH">
        <Focus="BULLFIGHT">
        <Focus="LEAGUE">
        <Focus="CUP">
        <Focus="RACE">
      <Focus="PRESENTATION">
      <Focus="POLITICAL_EVENT">
        <Focus="ELECTIONS">
      <Focus="MUSICAL_EVENT">
        <Focus="CONCERT">
        <Focus="FESTIVAL">
      <Focus="CONTEST">
        <Focus="FEW_PUBLIC_EXAMS">
      <Focus="ACT">
      <Focus="PRIZE">
      <Focus="TRIBUTE">
      <Focus="EXHIBITION">
      <Focus="COURSE">
  <Focus="PRODUCT">
    <Focus="ENTERTAINMENT">
      <Focus="PHOTOGRAPHY">
      <Focus="SONG">
      <Focus="MUSIC_VIDEO">
      <Focus="MOVIE">
      <Focus="MUSIC_ALBUM">
      <Focus="TV_SHOW">
        <Focus="TV_SERIES">
      <Focus="BOOK">
      <Focus="MAGAZINE">
      <Focus="PRESS">
      <Focus="VIDEOGAME">
    <Focus="TECHNOLOGY">
      <Focus="HARDWARE">
      <Focus="SOFTWARE">
        <Focus="INTERNET">
          <Focus="SOCIAL_NETWORK">
          <Focus="SEARCH_ENGINE">
      <Focus="MOTOR">
        <Focus="CAR">
        <Focus="MOTORBIKE">
  <Focus="SITUATION">
    <Focus="ECONOMICAL">
      <Focus="REAL-STATE_MARKET">
    <Focus="LABORAL">
      <Focus="UNEMPLOYMENT">
      <Focus="STRIKE">
      <Focus="EMPLOYMENT">
      <Focus="FINANCING">
        <Focus="SCHOLARSHIP">
    <Focus="EDUCATION">
  <Focus="DISCUSSION_TOPIC">
    <Focus="IMMIGRATION">
    <Focus="ABORTION">
    <Focus="GENDER_EQUALITY">
    <Focus="NAZISM">
      <Focus="NEONAZISM">
    <Focus="RACISM">
    <Focus="FEMINISM">
    <Focus="MALE_CHAUVINISM">
    <Focus="TURISM">
    <Focus="HOMOSEXUALITY">
    <Focus="SEPARATISM">
    <Focus="PROSTITUTION">
    <Focus="NARCOTRAFFIC">
    <Focus="CAPITALISM">
    <Focus="SOCIALISM">
    <Focus="NUCLEAR_ENERGY">
    <Focus="RENEWABLE_ENERGY">
</LexiconHierarchy>
```

References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the 3rd international workshop on link discovery* (pp. 36–43). Chicago, Illinois.

Agrawal, R., Rajagopalan, S., Srikant, R., & Xu, Y. (2003). Mining newsgroups using networks arising from social behavior. In *Proceedings of the 12th international conference on World Wide Web*. Hungary, Budapest.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2007). Show me the money!: deriving the pricing power of product features by mining consumer reviews. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. USA, San Jose, California.

Baroni, M., & Vegnaduzzo, S. (2004). Identifying subjective adjectives through web-based mutual information. In *Proceedings of the 7th German conference on natural language processing (KONVENS'04)* (pp. 613–619).

Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In OMNIPRESS (Ed.), *Proceedings of ICWSM 2007*. Boulder, Colorado.

Bickart, B., & Schindler, R. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing, 15*, 31–40.

Carenini, G., Ng, R., & Zwart, E. (2005). Extracting knowledge from evaluative text. In *Proceedings of the international conference on knowledge capture (K-CAP'05)* (pp. 11–18). Banff, Alberta, Canada.

Chen, H., & Zimbra, D. (2010). Ai and opinion mining. *IEEE Intelligent Systems, 25*, 74–80.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 345–354.

Choi, Y., Breck, E., & Cardie, C. (2006). Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 conference on empirical methods in natural language processing (EMNLP 2006)* (pp. 431–439). Sydney.

Choi, Y., & Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP '08)* (pp. 793–801). Stroudsburg, PA, USA.

Danescu-Niculescu-Mizil, C., & Lee, L. (2010). Dont have a clue? unsupervised co-learning of downward-entailing operators. In *Proceedings of the ACL short papers* (pp. 247–252).

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems, 25*.

Delort, J. (2006). Identifying commented passages of documents using implicit hyperlinks. In *Proceedings of HYPERTEXT'06* (pp. 89–98). Odense, Denmark.

Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. In *Proceedings of the IEEE 24th international conference on data engineering workshop (ICDEW 2008)* (pp. 507–512). IEEE Computer Society.

Denecke, K. (2009). Are sentiwordnet scores suited for multi-domain sentiment classification? In *International conference on digital information management*.

Devitt, A., & Ahmad, K. (2007). Sentiment polarity identification in financial news: A cohesion-based approach. In A. Press (Ed.), *Proceedings of the 45th annual meeting association* (pp. 984–991). Computational Linguistics.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the conference on web search and web data mining (WSDM'08)* (pp. 231–239).

Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledgemanagement*. Bremen, Germany.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th conference on language resources and evaluation (LREC)*. Genova, Italy.

Godbole, N., Srinivasaiah, M., & Skiena, S., 2007. Large-scale sentiment analysis for news and blogs. In *Proceedings of the international conference on weblogs and social media (ICWSM)*.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of 35th meeting of the association for computational linguistics (pp. 174–181)*. Madrid, Spain.

Hoffman, T. (2008). Online reputation management is hot – But is it ethical? Computerworld.

Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. USA, Seattle, WA.

Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. In *Proceedings of nineteenth national conference on artificial intellgience (AAAI'04)*.

Hu, M., Sun, A., & Lim, E. (2007). Comments-oriented blog summarization by sentence extraction. In *Proceedings of the sixteenth ACM conference on information and knowledge management*. Portugal, Lisbon.

Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. In *AAAI'06*.

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the conference on web search and web data mining (WSDM)* (pp. 219–230). Stanford, CA.

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. In *Proceedings of the international conference on computational linguistic (COLING)*. Geneva, Switzerland.

Kobayashi, N., Inui, K., & Matsumoto, Y. (2007). Extracting aspect-evaluation of aspect-of relations in opinion mining. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 1065–1074). Prague, Czech Republic.

Li, G., Feng, J., Wang, J., Yu, B., & He, Y. (2008). Race: Finding and ranking compact connected trees for keyword proximity search over xml documents. In *Proceedings of the 17th international world wide web conference (WWW 2008)* (pp. 1045–1046). Beijing, China.

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating classification and association rule mining. In *KDD'98*.

Lloyd, L., Mehler, A., & Skiena, S. (2006). Identifying co-referential names across large corpora. *Lecture Notes on Combinatorial Pattern Matching, 4009*, 12–23.

Lue, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: An optimization approach. In *Proceedings of the 20th international conference on world wide web* (pp. 347–356). Hyderabad, India.

Malouf, R., & Mullen, T. (2008). Taking sides: User classification for informal online political discourse. *Internet Research, 18*, 177–190.

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Massachusetts: In MIT Press.

McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse sentiment analysis. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 432–439). Prague, Czech Republic.

Miao, Q., Li, Q., & Dai, R. (2008). An integration strategy for mining product features and opinions. In *Proceedings of the 17th ACM conference on information and knowledge management* (pp. 1369–1370). Napa Valley, California.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography (special issue), 3*, 235–312.

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP* (pp. 412–418).

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*. USA, Sanibel Island, FL.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). Sentiful: A lexicon for sentiment analysis. *IEEE Transactions on Affective Computing, 2*, 22–36.

Ohana, B., & Tierney, B. (2009). Sentiment classification of reviews using sentiwordnet. In *9th. IT&T conference*.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*, 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (pp. 79–86). Philadelphia: Association for Computational Linguistics.

Popescu, A., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of EMNLP* (pp. 339–346).

Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international joint conferences on artificial intelligence* (pp. 1199–1204). Pasadena, California.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*, 267–307.

Tan, S., & Wang, Y. (2011). Weighted scl model for adaptation of sentiment classification. *Expert Systems with Applications, 38*, 10524–10531.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications, 38*, 12094–12100.

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications, 34*, 2622–2629.

Täckström, O., & McDonald, R. (2011). Semi-supervised latent variable models for sentence-level sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics* (pp. 569–574). Portland, Oregon.

Titov, I., & McDonald, R. (2008). A joint model of text and aspect ratings for sentiment summarization. In *The 46th annual meeting of the association for computational linguistic* (pp. 308–316). Columbus, Ohio.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*. Philadelphia, Pennsylvania.

Velikovich, L., Blair-Goldensohn, S., Hannan, K., & McDonald, R. (2010). The viability of web-derived polarity lexicons. In *The 2010 annual conference of the North American chapter of the ACL* (pp. 777–785). Los Angeles, California.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics, 30*, 277–308.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, 2*, 165–210.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing (HLT/EMNLP)* (pp. 347–354). Vancouver.

Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? finding strong and wake opinion clauses. In *Proceedings of AAAI* (pp. 761–769). San Jose, CA.

Wu, Q., & Tan, S. (2011). A two-stage framework for cross-domain sentiment classification. *Expert Systems with Applications, 38*, 14269–14275.

Yessenalina, A., Yue, Y., & Cardie, C. (2010). Multi-level structured models for document-level sentiment classification. In *Proceedings of the 2010 conference*

on empirical methods in natural language processing (pp. 1046–1056). Massachusetts, USA.

Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the third IEEE international conference on data mining*.

Zhai, Z., Xu, H., Kang, B., & Jia, P. (2011). Exploiting effective features for chinese sentiment classification. *Expert Systems with Applications, 38*, 9139–9146.

Zhang, Z., & Varadarajan, B. (2006). Utility scoring of product reviews. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. USA, Arlington, Virginia.